Package: RFclust (via r-universe)

October 29, 2024

Type Package

Title Random Forest Cluster Analysis

Version 0.1.2

Author Ankur Chakravarthy, PhD

Maintainer Ankur Chakravarthy <ankur.chakravarthy.10@ucl.ac.uk>

Description Tools to perform random forest consensus clustering of different data types. The package is designed to accept a list of matrices from different assays, typically from high-throughput molecular profiling so that class discovery may be jointly performed. For references, please see Tao Shi & Steve Horvath (2006) <doi:10.1198/106186006X94072> & Monti et al (2003) <doi:10.1023/A:1023949509487> .

License GPL

Encoding UTF-8

Imports ConsensusClusterPlus,randomForest

NeedsCompilation no

Depends R (>= 3.5.0)

Date/Publication 2022-06-21 02:30:02 UTC

Repository https://archakravarthy.r-universe.dev

RemoteUrl https://github.com/cran/RFclust

RemoteRef HEAD

RemoteSha 94bf809d3ead79a785b5dee087b88c1b3650d9ea

Contents

gbm *Multi-omic profiling of glioblastoma samples*

Description

These data serve as an example dataset to execute RFCluster on.These were processed and originally included in the iCluster R package which is likely to be archived on CRAN. I have reincluded this dataset here to permit the example to be run.

Usage

```
data("gbm")
```

Format

A list of matrices containing Copy Number, Methylation and Expression estimates for 55 GBMs for 1500-1800 genes. Data were originally derived by The Cancer Genome Atlas project.

Source

https://doi.org/10.1093/bioinformatics/btp659

Examples

```
data(gbm)
```

RFCluster *A wrapper for Random Forest Consensus Clustering*

Description

This takes a list of matrices of different data types , features in rows, samples in columns, and performs random forest clustering (one-dimensional). When multiple data types are available this is one way of modelling the data together.

Usage

```
RFCluster(Data, ClustAlg = "pam", MaxK, nTrees = 1000,
exportFigures = "pdf", ClustReps = 500, ProjectName = "RFCluster",
verbose = TRUE, ...)
```

Arguments

`Data` Named list, contains matrices with samples in columns, features in rows. The names of the list should represent the platform or the feature type, such as expression, or CN, or clin; as long as it is distinct.

`ClustAlg` Algorithm for consensus clustering

`MaxK` Maximum number of clusters you are searching for

`nTrees` How many trees are we using in the random forest to generate a proximity matrix?

`ProjectName` Name of the project, to annotate plots and other output

`ClustReps` Number of replicates for consensus clustering

`verbose` Should output be verbose?

`exportFigures` Format of the results file for figures et cetera to be exported to

`...` Other optional arguments, passed onto ConsensusClusterPlus; see that package's documentation for a full set.

Value

Standard output for ConsensusClusterPlus runs.

Author(s)

Ankur Chakravarthy, PhD

References

Monti, S., Tamayo, P., Mesirov, J. et al. Machine Learning (2003) 52: 91. https://doi.org/10.1023/A:1023949509487

Tao Shi & Steve Horvath (2006) Unsupervised Learning With Random Forest Predictors, Journal of Computational and Graphical Statistics, 15:1, 118-138, DOI: 10.1198/106186006X94072

Examples

```
library(RFclust)

#Get GBM example data from the iCluster package, repackaged to maintain CRAN compatibility
data(gbm)

#Transpose so columns are samples and features are rows
gbm.t <- lapply(gbm, t)

#Make sure the sample names are the same across the matrices for the different
#samples - the code breaks otherwise

colnames(gbm.t[[2]]) <- colnames(gbm.t[[3]]) <- colnames(gbm.t[[1]])

#Run function on that dataset - these methods are computationally intensive
#so automatic testing during build has been disabled (takes > 5s).
#Users may test the software by running the code separately as the example is reproducible
```

```
Test.cluster <- RFCluster(Data = gbm.t, ClustAlg = "pam", MaxK = 5,
nTrees = 10, ProjectName = "RFCluster_Test", ClustReps = 50 , writeTable = FALSE, plot = NULL)
unlink("RFCluster_Test",recursive = TRUE)
```

Index